

知识图谱在青铜器数字馆藏建设中的应用*

胡汗林 邓三鸿

(南京大学信息管理学院, 南京 210023)

摘要: 随着大数据时代的到来, 以博物馆为代表的各大文化与科研机构展开文物数字化和数字馆藏建设的研究。青铜器是中国古代礼仪制度的重要实体表现, 然而目前国内青铜器研究大多集中在学术文献与博物馆中, 存在文物资源信息分散、数据质量不佳、数据关联性不强等问题。针对这些问题, 对青铜器概念与术语进行分析, 构建青铜器本体模型, 将本体模型映射到图数据库中, 利用Neo4j数据库构建青铜器知识图谱并探讨相关应用, 实现对青铜器知识的关联与挖掘, 从而促进知识共享与发现, 为青铜器数字馆藏建设提供新思路。

关键词: 青铜器; 本体; 知识图谱; 数字馆藏; 可视化

中图分类号: G250.73; TP391.1 **DOI:** 10.3772/j.issn.1673-2286.2023.04.001

引文格式: 胡汗林, 邓三鸿. 知识图谱在青铜器数字馆藏建设中的应用[J]. 数字图书馆论坛, 2023 (4) : 1-8.

随着大数据时代的到来, 文物数字化、智能博物馆等概念逐渐走进人们的视线。2020年中宣部文改办发《关于做好国家文化大数据体系建设工作通知》, 明确提出推进文化和科技深度融合, 分类采集梳理文化遗产数据, 对全国公共文化机构、高等科研机构和文化生产机构各类藏品数据标注文化基因, 青铜器文物数据自然也是其中的重点梳理对象^[1]。2022年中共中央办公厅和国务院办公厅印发《关于推进实施国家文化数字化战略的意见》, 明确指出到“十四五”末, 基本建成文化数字化基础设施和服务平台^[2]。在国家的大力号召和相关新兴技术的快速发展下, 各文化与科研机构对数字馆藏的研究愈发深入, 越来越多的博物馆投入智慧博物馆的建设, 包括青铜器在内, 越来越多的文物资源通过种种数字化手段呈现在大众面前。

青铜器是中国古代礼仪制度的物质表现, 其独特的造型与装饰反映了商周时期的艺术成就, 青铜器铭文为了解古人的生活和思想提供了重要史料, 其成熟的工艺与铸造技术更为世界瞩目, 传递着中华民族的智慧^[3]。近些年, 随着出土文物数量的不断增加, 青铜器愈发受到国人的关注, 引起了广泛的研究。同时青铜器

是中国考古学和博物馆学中的重要门类, 已有的研究十分丰富, 大量的图录和研究成果出版。通过对其进行线上研究, 可以进一步拓宽其研究门类, 有助于进一步挖掘该领域的知识。

尽管国家极力促进文化遗产的数字化建设工作, 但青铜器数字馆藏建设中仍存在文物资源信息分散、数据质量不佳、数据关联性不强等问题, 目前缺乏对青铜器语义层面的研究。基于此, 本文对国内多个知名博物馆的馆藏青铜器数据进行集成, 分析概念与属性, 以统一的数据标准对青铜器属性进行规范, 对青铜器知识进行语义关联, 进而构建青铜器的本体模型和知识图谱, 为青铜器等实物资源的知识共享与传播提供新的思路, 为青铜器数字馆藏建设提供新方案。

1 相关研究现状

1.1 青铜器整理相关研究

青铜器作为中国传统文物的重要代表, 大多存放

收稿日期: 2023-03-22

* 本研究得到国家社会科学基金重点项目“大数据环境下领域知识加工与组织模式研究”(编号: 20ATQ006)资助。

在全国各地的博物馆中,目前有关青铜器的研究以博物馆和相关文化机构为主,研究成果多为图录书籍。

《中国青铜器图录(下)》^[4]是系统介绍青铜器相关知识和数据的图录,该书按朝代对青铜器的名称、时代、尺寸、出土、馆藏等类别数据进行介绍,此外还阐述了青铜器分类和辨伪等知识,是很好的科普性书籍,但该书并未列出青铜器目录,不利于用户快速查找。

针对青铜器信息分散的问题,科学出版社策划出版了《中国出土青铜器全集》^[5],对中国出土的青铜器作了全面、科学的梳理,分省成册。书籍中收录的青铜器图片精美,信息全面,同时对于每个青铜器都有目录索引,便于读者快速定位目标青铜器页。

《商周青铜器铭文暨图像集成索引》^[6]是《商周青铜器铭文暨图像集成》及其续编的配套索引书,它对于书中的族氏族徽、人名、地名、器物出土地、器物现藏地等分类设置了索引,读者可以根据想要检索的内容在检字表中快速定位描述页。

通过调研,可以看出目前青铜器相关书籍以展示青铜器相关属性和图片为主,由于青铜器文物数量较多,部分书籍会出版多册以对青铜器进行系统化整理和展示。此外,部分书籍还为其整理的青铜器设置目录或配套索引以帮助读者快速定位具体页码。然而,书籍不便于一般用户获取和翻阅,也不便于专业学者的研究。在大数据时代,更应该探讨对青铜器等文物进行数字化存储与展示的方式。

1.2 文物数字馆藏建设相关研究

随着大数据、云存储等技术的发展,文物数字化存储与展示已成为可能。2012—2016年我国开展了第一次全国可移动文物普查工作^[7],对我国可移动文物进行全面调查登记并建立了全国可移动文物信息登录平台和数据库。除了国家层面的重大国情国力调查工作以外,国内外各大博物馆和相关学者也积极探索如何对馆藏文物更好地进行数字化存储与系统性展示。

Frenger^[8]对美国新墨西哥州的Georgia O'Keeffe Museum推出的藏品网站进行了详细介绍。该网站以关联数据的形式对藏品进行数字化存储,每个藏品的详情页列出与该藏品同年代生产或出现在同一展览上的其他藏品,为用户深入、系统地了解相关藏品知识提供了有利条件。Petras等^[9]与Meghini等^[10]将欧洲大型数字文化遗产搜索引擎Europeana作为研究对象。

Europeana为欧亚大陆不同国家/地区的数字文化遗产项目提供了元数据描述的单一访问节点,使用了汇集多个不同文化遗产标准的数据模型EDM(Entity Data Model),在EDM中引入事件的概念,通过对资源的描述,引出更多相关资源,体现了数字资源的关联性。

张前军^[11]基于语义网为中国青铜器建立本体库,对收集的青铜器文物进行语义标注并建立检索模型,实现了青铜器领域智能检索系统,但该系统体量较小,系统响应时间较长,本体的可扩展性较差,不适合大规模推广。李峰^[12]针对目前国内数字馆藏建设过于零散化和个体化、文博界缺乏大规模系统性知识组织与传播的问题,指出采用更多维度、多层次和细粒度的知识组织方式构建文物本体和知识图谱,从而实现知识互通。李婉娜等^[13]等基于文化基因的相似性,对革命文物进行关系匹配,构建革命文物的数字化聚合体,为革命文物数字化保护提供了新思路。高劲松等^[14]针对馆藏文物资源异构、种类繁多、关联程度低等问题,提出基于关联数据的馆藏文物知识关联方法,构建知识关联模型,利用关联数据技术实现对馆藏文物资源的访问和查询。冯秋燕等^[15]将元数据、图数据库等相关技术融入文博机构文物数字馆藏的建设,构建适用于可移动文物的本体框架,并以铜器为实例证明了该框架的可行性。

综合来看,国外的文物数字馆藏建设较为成功,有良好的实践案例,在数据规模上已经实现资源聚合与集中展示,通过数据模型实现了数字资源间的关联。国内对于文物数字化和智能博物馆的建设工作较为零散,存在数据孤岛问题,数据资源关联程度低,大多数项目都处于计划阶段,资源整合与实践落地不足。

1.3 文物类知识图谱构建进展

知识图谱是谷歌公司于2012年提出的结构化语义库,主要用于描述物理世界中的概念及其相互关系,它以图的形式将实体与实体间的关系进行可视化呈现。按功能和使用场景,知识图谱可以分为以搜索引擎为主要形式的通用知识图谱和面向具体领域的领域知识图谱,领域知识图谱相较于通用知识图谱更强调知识的深度,知识结构更复杂,知识质量要求更高^[16]。目前,领域知识图谱已经被广泛运用于金融^[17]、医疗^[18]、政务^[19]与公安^[20]等行业的数字化建设中,通过出色的知识组织与推理能力为这些行业的核心业务场景提供技术支持。在文博领域,知识图谱同样可以发挥重要的知识组

织作用: 通过为文物类数据构建领域知识图谱, 可以直观呈现出文物实体间的关联关系, 挖掘出实体间的隐性链接, 为文物数字馆藏建设提供有力支持。

高劲松等^[21]通过对文物领域概念的分析并部分复用CIDOC CRM模型, 构建了包含5个核心类、4个子类、18个对象属性和22个数据属性的可移动文物本体模型, 随后基于该本体, 利用Neo4j数据库对湖北省博物馆的部分文物数据进行知识图谱构建, 实现了对文物本体的多维度分析。李永卉等^[22]以南朝陵墓石刻为研究对象, 借鉴七步法与循环法构建石刻本体模型, 以丹阳陵墓石刻为例, 根据石刻本体模型, 利用Neo4j数据库构建了南朝陵墓石刻知识图谱。张娜^[23]针对国内博物馆在构建知识图谱时自动化程度较低的问题, 对文物关系自动抽取技术进行研究, 最终实现了一个支持各类专题文物知识图谱的文物知识图谱构建方案与系统。彭博^[24]通过分析网络文物信息资源的内容与结构特征, 利用关键词与外部知识库进行文物知识抽取并构建知识图谱, 为文物信息资源的深度开发提供参考。为了缓解文物领域人工标注数据匮乏的问题, Zhang等^[25]提出基于变压器的双向编码表示的实体类型信息模型, 用于文物文本知识图谱补全, 该模型能丰富实体的语义信息, 在三元组分类、链接预测和关系预测等任务中取得了比基线模型更好的性能。

综合来看, 近些年, 国内学者已经开始尝试将知识图谱技术应用到文物数字馆藏建设中, 研究重点为图谱构建与模型创新。然而, 目前大多数研究更关注整个文物大类的知识图谱构建, 泛化性较强, 缺乏对某一类文物特性及相关知识的运用, 对特定类别文物领域的专业性研究存在不足, 难以挖掘出该领域特有的隐性知识。为了深度挖掘青铜器领域的隐性知识, 将知识图谱技术融入青铜器数字馆藏建设, 通过收集与调研其数据资源, 构建青铜器领域的本体模型, 利用图数据库对数据资源进行存储, 并以知识图谱的形式发布青铜器实例数据, 对青铜器资源的关联关系进行可视化呈现, 将隐性知识显性化, 从而为青铜器数字馆藏建设提供新方案。

2 青铜器本体模型构建

本体^[26]是通过描述和捕捉领域知识来发现领域中的概念与概念之间的关系, 使领域内不同的主体之间可以进行沟通和知识共享的形式化规范说明。常用的

领域知识本体构建方法有METHONTOLOGY法^[27]、SENSUS法^[28]、七步法^[29]等。参考最常用的七步法, 构建青铜器领域本体。

2.1 青铜器核心概念与术语

青铜器大多存放于博物馆中, 青铜器数据也大多呈现在各博物馆的官网中。对北京故宫博物院^[30]、河南博物院^[31]、安徽博物院^[32]、台北故宫博物院^[33]等国内知名博物馆的网站进行调研, 辅以专业图录书籍, 对青铜器的属性进行较为全面的收集与整理。

在数据采集的过程中, 由于博物馆官网的青铜器数据结构化程度较高, 在利用爬虫工具采集后, 获取的青铜器“名称”“尺寸”“时代”“收藏单位”等大多数青铜器属性可直接作为数据集中的字段, 而例如“纹饰”“出土地”“出土时间”等字段则均包含在“资料”属性中, 通过Excel软件的拆分列与字符串匹配等功能, 将这些字段从“资料”属性中抽取出来, 再辅以人工检查, 筛选得到较为准确的字段与值。整理出的字段名为青铜器领域概念, 字段值则为相应概念的术语。整理后的青铜器领域部分重要概念与术语如表1所示。

表1 青铜器领域部分概念与术语

概念	术语
形制	鼎、尊、甗、鬲、簋、豆、盃、簠、爵、角、觚、觥、罍、觥……
类别	酒器、食器、水器、兵器、量器……
时代	夏、商代早期、商代晚期、西周早期、西周晚期、春秋早期……
纹饰	兽面纹、云雷纹、乳钉纹、蕉叶纹、夔纹、蟠虺纹、饕餮纹……
收藏单位	北京故宫博物院、河南博物院、安徽博物院、台北故宫博物院……
出土地	陕西宝鸡、河南新郑、河南安阳、山东益都、安徽省寿县……

2.2 本体模型核心类

在明确青铜器的核心概念和属性后, 参考其他文物类知识图谱的本体概念层级^[21], 并根据所采集到的青铜器属性特征, 尝试构建青铜器概念体系。作为青铜器领域本体模型, 大多字段属于青铜器的数据属性, 因此须划分“青铜器”大类。此外, “类别”作为与青铜

器高度相关且十分重要的字段,在青铜器知识体系中具有重要地位,因此将其单独列为一个大类与青铜器进行关联。“时间”“地点”“资料”等字段具有独特含义,在知识体系中同样具备重要作用,不宜直接作为青铜器类的数据属性,因此将它们分别划分为3个大类。最终,构建的青铜器概念体系包含5个核心类,分别为“青铜器”“类别”“时间”“地点”“资料”。

(1) 青铜器。青铜器类是指青铜器的实体,包含编号、名称、质量、高度、尺寸、形制、纹饰等有关青铜器的基本描述性信息。编号为构建的本体中青铜器的唯一标识符,质量和高度是统一单位后的数值型数据,统一单位为“g”和“cm”。由于部分青铜器形状特殊,除了简单的高度和质量描述,还需要详细的文本型尺寸信息描述数据,如“错金银鸟耳壶”的尺寸描述是“高36.9cm,口径17.4cm,腹径26.1cm,足径13.9cm,重5 880g”,对于这种壶形文物,口径、腹径、足径也是重要的尺寸参数。形制是青铜器的造型信息,通常与青铜器名称的最后一个字有关,如“后母戊鼎”的最后一个字“鼎”能代表其形制。纹饰是青铜器表面的纹样,一件青铜器表面通常有较多纹饰,因此纹饰是多值型文本数据。

(2) 类别。类别类是指青铜器的类别,通常也可视作青铜器的用途。青铜器的类别一般与其形制有关,现有青铜器分类标准较多。参考《中国青铜器图录(下)》^[4]的分类标准,辅以百科知识,将青铜器分为7类:食器、酒器、水器、乐器、兵器、量器、杂器。其中,食器和酒器还可以进行更细粒度的划分,但部分形制的类别不唯一(如尊既可以是饪酒器,也可以是储酒器),为了使类别的层级统一,不作细分,取最高类别层级。

(3) 时间。时间类是指青铜器数据中的时间信息,包含朝代和出土时间两个子类。青铜器产生的朝代主要为夏朝到汉朝,部分青铜器的朝代信息显示为早期、中期、晚期或前期、中期、后期,在此统一为早期、中期、晚期。出土时间为青铜器出土的具体年份信息。

(4) 地点。地点类是指青铜器数据中的地点信息,包含收藏单位和出土地点两个子类。青铜器的收藏单位主要为全国各地的博物馆。出土地点子类包含出土地和墓葬两个属性,出土地为省、市、县等行政区域信息的组合,墓葬为具体的墓葬名称。

(5) 资料。资料类是青铜器其他的描述性信息,主要包含器物介绍、历史价值和图片等信息。器物介绍是对青铜器外形的详细描述;历史价值为青铜器在历

史上的价值,多为考古价值,可以从中发掘历史故事或该件青铜器的历史意义;图片即青铜器的图片,以可直接访问的URL形式存储。

2.3 对象属性与数据属性

在本体中,最基础的元素是三元组(SPO),本体模型中的类是三元组中的S和O,在明确本体的核心类与子类后,需要创建相应的属性(P)。本体中的属性分为对象属性(object properties)和数据属性(data properties),对象属性用来定义类之间的关系,数据属性用来定义类具有的属性。青铜器本体模型的对象属性和数据属性见表2和表3。

表2 青铜器本体对象属性

对象属性	含义	定义域(domain)	值域(range)
belongTo	属于	青铜器	类别
collected	藏有	收藏单位	青铜器
excavatedIn	出土于	青铜器	出土时间
hasData	有资料	青铜器	资料
isCollectedIn	收藏于	青铜器	收藏单位
producedIn	产生于	青铜器	朝代
timeIs	时间是	出土地点	出土时间
unearthed	出土了	出土地点	青铜器
unearthedIn	出土在	青铜器	出土地点

表3 青铜器本体数据属性

定义域(domain)	数据属性	值域(range)
青铜器	名称	xsd:string
	尺寸	
	形制	
	纹饰	
	编号	xsd:int
	质量	
类别	类别	xsd:string
朝代	朝代	xsd:string
出土时间	出土年份	xsd:string
收藏单位	单位名称	xsd:string
出土地点	出土地	xsd:string
	墓葬	
资料	器物介绍	xsd:string
	历史价值	
	图片	xsd:anyURI

在定义对象属性时,可以定义属性之间的互逆关系,这有助于优化本体模型的推理关系。在构建的青铜器本体模型中,存在两对互逆属性:第一对是unearthed与unearthedIn,其定义域和值域为青铜器和出土地点;第二对是collected与isCollectedIn,其定义域与值域为青铜器和收藏单位。

2.4 本体模型绘制

在定义青铜器本体模型的类、对象属性和数据属性后,可通过本体开发工具Protégé绘制该本体模型,最终构建出包含5个核心类、4个子类、9个对象属性、16个数据属性的青铜器本体模型。

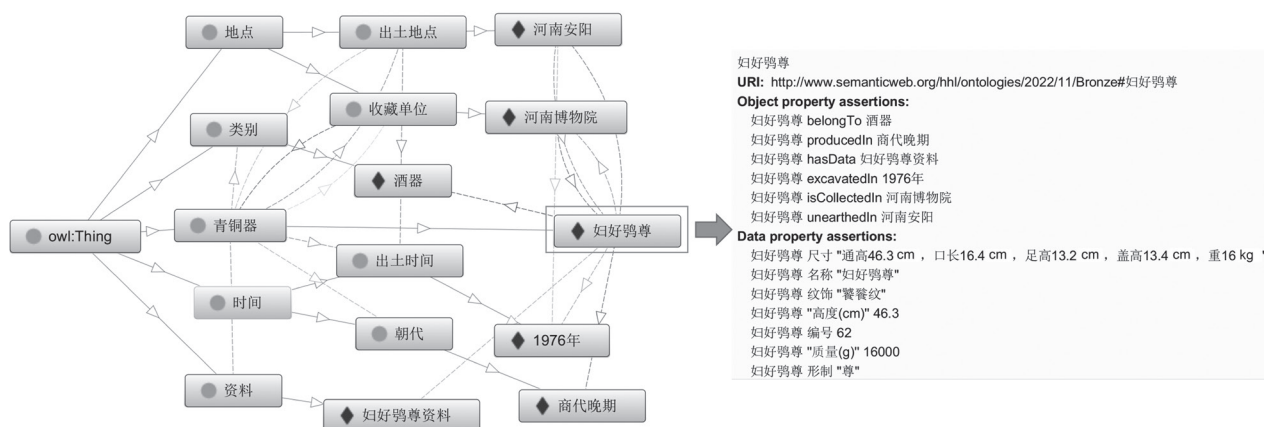


图1 青铜器本体实例图

3 青铜器知识图谱与可视化

知识图谱最早用来提升搜索引擎的能力,随后,知识图谱在辅助智能问答、自然语言理解、大数据分析、推荐计算等方面都展现出丰富的应用价值^[16]。知识图谱按功能和应用场景分为通用知识图谱和领域知识图谱,针对青铜器数字馆藏建设,所要构建的是青铜器领域知识图谱。相较于搜索引擎使用的通用知识图谱,其知识结构更加复杂,知识抽取质量更高,通常需要提前构建领域本体。

3.1 构建工具与数据源

使用的知识图谱构建工具为Neo4j数据库,Neo4j数据主要有节点、标签、关系、属性4类构建元素,这4类元素与本体模型中的元素一一对应。Neo4j数据库中

基于已构建的青铜器本体模型,以妇好鸮尊为例,将该青铜器的实例数据应用于本体模型,以验证青铜器本体模型的可行性与实用性,可视化结果如图1所示,其中圆圈节点表示概念类,菱形节点表示实例实体。可以看出,妇好鸮尊作为青铜器实例,朝代为商代晚期,出土时间为1976年,出土地点为河南安阳,类别为酒器,收藏单位为河南博物院,相关资料整合在一个实体中。同时,右侧的详情框展示了妇好鸮尊的统一资源标识符 (URI)、对象属性和数据属性等详细信息。通过实例化展示可见,已构建的本体模型可以对青铜器数据进行较好的组织,可以依据该本体模型进一步构建青铜器知识图谱。

的节点与本体中的实例对应,标签与本体中的类对应,关系与本体中的对象属性对应,属性与本体中的数据属性对应。通过这4个对应关系,可以将本体模型映射到Neo4j数据库中。

青铜器主要收藏于全国各大博物馆中,其较为全面和权威的数据也来自各大博物馆。选取国内多个知名博物馆,对其官网展示的青铜器数据进行采集,最终从12个博物馆官网上采集了137条青铜器数据并导入Excel软件,根据已构建的青铜器本体模型对数据属性进行数据清洗工作,最终形成csv格式的数据集文件。

3.2 青铜器知识图谱构建与应用

Neo4j数据库的查询语句是Cypher语句,通过Cypher语句可以将csv格式的数据导入Neo4j数据库从而形成实体和关系。最终形成的青铜器知识图谱包含7

个实体标签、372个实体、9个关系标签和1 136个关系。

构建出的青铜器知识图谱可以对青铜器资源的关联关系进行直观呈现,各实体间通过谓词关系相互链接。基于构建的青铜器本体模型,可以实现青铜器相关数据的自动补充与实例添加,快速与已有实体形成关联,充分发挥图存储的优势。

在实际应用中,该青铜器知识图谱可以在智能搜索、知识问答、个性化推荐、学术研究、数字化展览、知识发现等方面发挥重要作用。

(1) 智能搜索。知识图谱起初就是谷歌为了提高搜索质量而提出的,因此青铜器知识图谱可以对文博领域检索系统进行优化,实现智能搜索。具体而言,基于知识图谱的搜索引擎不再局限于输入语句的字面意思,而是可以通过自然语言处理技术理解输入语句的语义,找出语句中的实体和关系,提取关键词并将其转换为Cypher语句,实现在Neo4j数据库中的查询匹配。此外,由于知识图谱实现了对知识的关联,在进行智能化检索时不仅会返回准确的检索结果,还会返回与该结果实体相关的其他实体信息,如当用户搜索“妇好鸮尊”时,系统还会展现其他商代晚期青铜器的导航链接,为用户提供深度搜索与系统性学习相关知识的条件。

(2) 知识问答。知识问答系统可以依据用户的自然语言提问返回相关结果,青铜器领域知识图谱可以为博物馆相关机构的问答系统提供丰富的知识库,大大提高系统的回答质量和效率。当用户发出有关青铜器文物的提问时,系统可以识别提问者的真实意图,提取关键词,基于知识图谱中的三元组知识结构返回答案。例如,当用户提问“妇好鸮尊出土于什么时候?”,青铜器知识图谱可以提取出实体“妇好鸮尊”和关系“出土于”,基于图谱中的三元组数据返回结果“1976年”。

(3) 个性化推荐。个性化推荐系统是根据用户兴趣特征主动给用户可能感兴趣的信息的系统,这避免了传统方式下用户只能自主搜寻信息的弊端。博物馆在网站建设中,可以根据用户的属性信息和搜索历史、浏览记录等行为信息,基于知识图谱和推荐算法给用户推荐可能感兴趣的其他文物信息,让用户可以深入学习感兴趣领域的文物知识,提高博物馆信息服务的质量,提升用户对于博物馆网站的黏性^[34]。

(4) 学术研究。由于国内博物馆文物数据过于分散,相关学者想要系统性研究文物,难免会遇到困难,尤其是文史、考古等相关领域的学者,可能在其专业领

域有深刻造诣,但未掌握知识组织技术而难以实现对多源异构信息的分析。知识图谱作为知识组织工具,可以实现对文物数据的组织与存储,将文物知识进行关联,为文博领域相关学者的学术研究和交流提供知识库和平台保障。

(5) 数字化展览。随着信息技术的发展,各大博物馆将多种数字化技术融入线下展览活动,从多个维度给参展用户提供丰富的视觉体验。知识图谱作为优秀的知识组织与存储工具,可通过图片直观地呈现文物信息的关联关系,为博物馆的数字化展览提供更丰富的呈现方式。

(6) 知识发现。知识图谱通常通过图数据库的形式进行存储,如构建的青铜器知识图谱所用的Neo4j数据库,其内置了中心性、社群发现、相似度、路径寻找、链接预测等算法,这些算法可以支持对青铜器隐性知识的多角度挖掘,促进知识共享与发现。

3.3 可视化平台构建

本体建模与知识图谱构建,突破了传统模式数据存储与描述的局限,实现对青铜器知识语义层面的关联、存储与挖掘。为了让一般用户可以学习到更多青铜器知识,还依托Power BI工具构建了一个青铜器可视化平台,给用户提供了系统化学习青铜器文化的窗口,为青铜器的数字化推广提供了新方案。

以xlsx格式的数据集文件为该可视化平台数据源,通过Power BI工具中的多种视觉组件设计前端功能,最终构建了4个前端页面,分别为信息卡片页、分类展示页、动态图表页和出土分布页。其中,信息卡片页以卡片形式展示青铜器的图片和部分属性,页面上端提供青铜器名称检索功能,用户点击卡片即可看到青铜器的更多介绍性信息(见图2)。分类展示页以表格和图片的形式呈现青铜器的部分外形属性,同时提供对时代、类别、性质等字段的筛选功能。动态图表页以条形图、环形图、饼图等形式对青铜器的类别、时代、高度、质量等属性进行概览式呈现,让用户可以直观了解系统中青铜器的分布情况。该页面图表相互关联,用户点击一个图表中的某个元素,其余图表会相应发生动态改变。出土分布页主要以地图的形式对青铜器的出土地点分布情况进行展示,此外还提供收藏单位筛选功能,使用户可以直观了解不同博物馆馆藏青铜器的出土地点分布情况。



图2 信息卡片页

4 结语

本文从文物数字化的大背景出发, 针对目前青铜器数字馆藏建设中存在的文物资源信息分散、数据质量不佳、数据关联性不强等问题, 提出构建青铜器领域知识图谱, 重点阐述青铜器知识图谱的构建方式和应用。具体而言, 首先通过对青铜器领域概念和术语的分析, 利用Protégé工具创建了青铜器本体模型。随后, 将本体模型映射到图数据库中, 利用Neo4j数据库导入国内多个博物馆的青铜器数据, 完成青铜器知识图谱构建。最后, 利用Power BI工具构建面向普通用户的可视化平台, 为用户提供学习和了解青铜文化的窗口。

对青铜器数据资源进行整合, 构建的青铜器知识图谱不仅可以直观呈现数据间的关联关系, 还能对青铜器隐性知识进行挖掘, 有利于青铜器资源知识共享与发现, 为青铜器数字馆藏建设提供参考。然而, 构建的知识图谱体量较小, 缺乏足够的支撑, 在知识挖掘与推理上作用有限, 未来可以考虑扩充数据量以作进一步研究。

参考文献

- [1] 谭宇彤. 数字人文视域下档案众包项目公众参与意愿影响因素研究 [D]. 长春: 吉林大学, 2021.
- [2] 中共中央办公厅 国务院办公厅印发《关于推进实施国家文化数字化战略的意见》[EB/OL]. [2022-12-21]. <http://www.gov.cn/>
- [3] 李茜, 雷英. 《中国出土青铜器全集》(二十卷) 吉金重宝传递中华智慧 [N]. 中国新闻出版广电报, 2021-09-10 (T10).
- [4] 牛建伟, 牛瑞红. 中国青铜器图录 下 [M]. 北京: 中国商业出版社, 2000.
- [5] 李伯谦. 中国出土青铜器全集-8-安徽 [M]. 北京: 科学出版社, 2018.
- [6] 吴镇烽. 商周青铜器铭文暨图像集成索引 [M]. 上海: 上海古籍出版社, 2019.
- [7] 国家文物局. 第一次全国可移动文物普查工作报告 [R/OL]. [2022-12-06]. http://www.ncha.gov.cn/art/2017/4/7/art_722_139372.html.
- [8] FRENGER C. Georgia O' Keeffe museum: collections online [J]. Reference Reviews, 2011, 25 (1): 50-51.
- [9] PETRAS V, HILL T, STILLER J, et al. Europeana: a search engine for digitised cultural heritage material [J]. Datenbank-Spektrum, 2017, 17 (1): 41-46.
- [10] MEGHINI C, BARTALESI V, METILLI D, et al. Introducing narratives in Europeana: a case study [J]. International Journal of Applied Mathematics and Computer Science, 2019, 29 (1): 7-16.
- [11] 张前军. 基于语义网的面向青铜器领域检索研究 [D]. 西安: 西安理工大学, 2010.
- [12] 李峰. 文物知识聚合与传播的初步研究: 以上海博物馆“宋徽宗与他的时代数字人文专题”为例 [J]. 东南文化, 2022 (3): 169-177.
- [13] 李婉娜, 季铁, 王梓溪, 等. 革命文物数字化聚合体的构建与设

- 计应用研究[J]. 包装工程, 2022, 43(20): 146-153.
- [14] 高劲松, 方晓印, 刘思洋, 等. 基于关联数据的馆藏文物资源知识关联与智能问答研究[J]. 情报科学, 2021, 39(5): 12-20.
- [15] 冯秋燕, 李川, 朱学芳. 面向智能服务的数字馆藏资源组织与存储研究: 以文博机构为例[J]. 图书馆论坛, 2021, 41(9): 58-67.
- [16] 王昊奋, 漆桂林, 陈华钧. 知识图谱: 方法、实践与应用[M]. 北京: 电子工业出版社, 2019.
- [17] 吕华揆, 洪亮, 马费成. 金融股权知识图谱构建与应用[J]. 数据分析与知识发现, 2020, 4(5): 27-37.
- [18] 李贺, 刘嘉宇, 李世钰, 等. 基于疾病知识图谱的自动问答系统优化研究[J]. 数据分析与知识发现, 2021, 5(5): 115-126.
- [19] 华斌, 康月, 范林昊. 政策文本的知识建模与关联问答研究[J]. 数据分析与知识发现, 2022(11): 79-92.
- [20] 黄茜茜, 杨建林. 基于司法判决书的知识图谱构建与知识服务应用分析[J]. 情报科学, 2022, 40(2): 133-140.
- [21] 高劲松, 张强, 李帅珂. 可移动文物的知识图谱构建及关联数据存储: 以湖北省博物馆为例[J]. 现代情报, 2022, 42(4): 88-98.
- [22] 李永卉, 刘沁芑, 卢章平. 基于知识图谱的南朝陵墓石刻信息资源开发研究[J/OL]. 图书馆杂志: 1-19 [2022-12-21]. <https://kns.cnki.net/kcms/detail/31.1108.G2.20220928.1850.006.html>.
- [23] 张娜. 文物知识图谱构建关键技术研究与应用[D]. 杭州: 浙江大学, 2019.
- [24] 彭博. 数字人文视角下的网络文物信息资源知识图谱构建[J/OL]. 图书馆论坛: 1-11 [2023-01-28]. <http://kns.cnki.net/kcms/detail/44.1306.g2.20220317.1011.002.html>.
- [25] ZHANG M, GENG G H, ZENG S, et al. Knowledge graph completion for the Chinese text of cultural relics based on bidirectional encoder representations from transformers with entity-type information[J]. Entropy, 2020, 22(10): 1168.
- [26] 徐坚, 姚贤明, 王俊. 面向文本的领域本体构建中信息抽取技术研究[M]. 北京: 科学出版社, 2020.
- [27] FERNÁNDEZ-LÓPEZ M, GÓMEZ-PÉREZ A, JURISTO N. Methontology: from ontological art towards ontological engineering[C]//Proceedings of the AAAI97 Spring Symposium on Ontological Engineering. Stanford University, 1997: 33-40.
- [28] 李景. 本体理论及在农业文献检索系统中的应用研究: 以花卉学本体建模为例[D]. 北京: 中国科学院研究生院(文献情报中心), 2004.
- [29] GUARINO N, GIARETTA P. Ontologies and knowledge bases: towards a terminological clarification[M]//MARS N J I. Towards Very Large Knowledge Bases. Amsterdam: IOS Press, 1995: 25-32.
- [30] 故宫博物院[EB/OL]. [2022-12-22]. <https://www.dpm.org.cn/>.
- [31] 河南博物院[EB/OL]. [2022-12-22]. <http://www.chnmus.net/>.
- [32] 安徽博物院[EB/OL]. [2022-12-22]. <https://www.ahm.cn/>.
- [33] 台北故宫博物院[EB/OL]. [2022-12-22]. <https://www.npm.gov.tw/>.
- [34] 刘绍南, 杨鸿波, 侯霞. 文物知识图谱的构建与应用探讨[J]. 中国博物馆, 2019, 36(4): 118-125.

作者简介

胡汗林, 男, 2000年生, 硕士研究生, 研究方向: 知识图谱。

邓三鸿, 男, 1975年生, 博士, 教授, 通信作者, 研究方向: 知识图谱、科学计量, E-mail: sanhong@nju.edu.cn。

Application of Knowledge Graph in the Construction of Bronze Digital Collection

HU HanLin DENG SanHong

(School of Information Management, Nanjing University, Nanjing 210023, P. R. China)

Abstract: With the advent of the era of big data, major cultural and scientific research institutions represented by museums have carried out research on the digitization of cultural relics and the construction of digital collections. Bronzes are important entity of the etiquette system in ancient China. However, at present, most of the research on bronzes in China is concentrated in academic literature and museums, which has problems such as scattered information of cultural relic resources, poor data quality, and weak data correlation. Aiming at these problems, we analyze the bronze concepts and terms, construct the bronze ontology model, and map the ontology model to the graph database. Then, we use Neo4j database to build bronze knowledge graph and discuss related applications, realize the correlation and mining of bronze knowledge, so as to promote knowledge sharing and discovery and provide new ideas for the construction of bronze digital collection.

Keywords: Bronze; Ontology; Knowledge Graph; Digital Collection; Visualization

(责任编辑: 雷雪)